Catalogue no. 82-003-X
ISSN 1209-1367

Health Reports

Linking the Canadian Community Health Survey and the Canadian Mortality Database: An enhanced data source for the study of mortality

by Claudia Sanmartin, Yves Decady, Richard Trudeau, Abel Dasylva, Michael Tjepkema, Philippe Finès, Rick Burnett, Nancy Ross and Douglas G. Manuel

Release date: December 21, 2016

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2016

All rights reserved. Use of this publication is governed by the Statistics Canada Open Licence Agreement.

An HTML version **is also available.**

Cette publication est aussi disponible en français.

Health Reports, Vol. 27, no. 12, pp. 10-18, December 2016 • Statistics Canada, Catalogue no. 82-003-X

Linking the Canadian Community Health Survey and the Canadian Mortality Database: An enhanced data source for the study of mortality

by Claudia Sanmartin, Yves Decady, Richard Trudeau, Abel Dasylva, Michael Tjepkema, Philippe Finès, Rick Burnett, Nancy Ross and Douglas G. Manuel

Abstract

Background: This study summarizes the linkage of the Canadian Community Health Survey (CCHS) and the Canadian Mortality Database (CMDB), which was performed to examine relationships between social determinants, health behaviours and mortality in the household population.

Data and methods: The 2000/2001-to-2011 Canadian Community Health Surveys were linked to the 2000-to-2011 CMDB using probabilistic methods based on common identifiers (names, date of birth, postal code and sex) for eligible respondents (85%; n = 614,774). Mortality records from January 1, 2000 through December 31, 2011 for people aged 12 or older were eligible for linkage (n = 2.774 million). The linkage was enhanced with information from the Historical Tax Summary File. Quality assessment consisted of internal and external validation. Cox survival analysis (age-adjusted) was conducted to estimate hazard ratios (HRs) associated with selected health behaviours.

Results: Overall, 5.3% of eligible CCHS respondents linked to a mortality record; false positive and false negative rates were 0.04% and 2.43%, respectively. Linkage rates were higher among males (5.8%) and people aged 75 or older (20.2%), reflecting known mortality risks. Survival analyses confirmed elevated mortality risk associated with heavy (HR 2.36, CI 1.84, 2.89) and light smoking (HR 1.91, CI 1.52, 2.33), compared with not smoking; underweight and obesity, compared with normal and overweight; low fruit and vegetable consumption; and lack of physical activity.

Interpretation: Linking health behaviour information from the CCHS to mortality data from the CMDB allows for a greater understanding of modifiable determinants of mortality.

Key words: Data linkage, health survey, mortality, risk factors

In most industrialized countries, vital statistics registries and national health surveys are cornerstones of health surveillance.

Mortality data compiled by vital statistics registries for administrative purposes can be tabulated by basic demographic characteristics (age and sex), province and cause of death. However, little is known about the socioeconomic, cultural or linguistic characteristics of those who die, or about the contributions of lifestyle and social factors to mortality risk.

Health surveys, by contrast, collect information about the health status and health behaviours of individuals, as well as their socioeconomic and cultural characteristics. However, these surveys are typically cross-sectional; no information is provided about respondent health beyond the date of the survey.

When complementary health survey and administrative data are combined through record linkage,[1-3] relationships between social determinants and health outcomes such as death can be analyzed in more depth.[4-8] For example, in Canada, mortality data have been linked to census results to examine differences in the risk of death across socioeconomic groups.[9,10] These linked data have also been used to calculate mortality rates for subpopulations such as immigrants and Aboriginal groups.[11-13] Similarly, researchers in other countries have used linked data to investigate mortality among immigrants, Indigenous peoples and prisoners.[14-17] A landmark Ontario study used linked health survey and mortality data to explore associations with smoking, poor diet, physical inactivity and high stress. Results demonstrated that 60% of Ontario deaths were attributable to these risk factors.[18]

The linkage described in the present study combines information from the Canadian Community Health Survey with mortality data from the Canadian Mortality Database. This article explains the record linkage process and presents results about associations between health behaviours and mortality among a representative sample of Canadians.

Data and methods

Canadian Community Health Survey

The cross-sectional Canadian Community Health Survey (CCHS) collects information about the health, health behaviours and health care use of the non-institutionalized household population aged 12 or older. The survey excludes full-time members of the Canadian Forces and residents of reserves and some remote areas, together representing about 4% of the target population. The CCHS was first conducted in 2000/2001 (cycle 1), and again in 2003 (cycle 2) and 2005 (cycle 3), each time with a sample of size of approximately 130,000. Starting in 2007, the survey was conducted annually (sample size of 65,000). Response rates ranged from 69.8% to 78.9%. Details about the sampling strategy and content are available elsewhere (www.statcan.gc.ca).[19]

Authors: Claudia Sanmartin (Claudia.sanmartin@canada.ca), Michael Tjepkema, Philippe Finès and Douglas G. Manuel are with the Health Analysis Division, Yves Decady is with the Health Statistics Division, Richard Trudeau is with the Special Surveys Division, and Abel Dasylva is with the Household Survey Methods Division at Statistics Canada. Rick Burnett is with Health Canada. Nancy Ross is with McGill University. Douglas G. Manuel is also with the Ottawa Hospital Research Institute.

The CCHS records eligible for linkage to the Canadian Mortality Database were respondents who consented to share their survey information with provincial and federal ministries of health and to link their responses to administrative data. Overall, 85.3% (n = 614,774) of CCHS respondents were eligible for linkage (Table 1). Special sampling weights were created for the linked CCHS data to adjust for those who did not agree to share and link.

Canadian Mortality Database

The Canadian Mortality Database (CMDB) is a census of all deaths registered in Canada. Deaths are reported by the provincial and territorial Vital Statistics Registries to Statistics Canada; the information provided includes cause and date of death, names, date of birth, and postal code at the time of death. Cause-of-death information is coded using the version of the *International Classification of Diseases* (ICD) in effect at the time of death. CMDB records eligible for linkage were deaths of people aged 12 or older (n = 2.774 million) that occurred from January 1, 2000 through December 31, 2011.

Historical Tax Summary File

The Historical Tax Summary File (HTSF) is a compilation of annual tax return files representing unique individuals for whom a tax declaration was produced in a given year. The HTSF contains only personal identifier information (names, date of birth, sex and postal code) from the T1 Income Tax Personal Master File; the file does not include income data. Statistics Canada uses the HTSF to assist in record linkage, specifically, through the provision of additional linkage information (names, postal code), and in manual resolution of doubtful links.

HTSF data for 1996 through 2011, representing about 33.5 million records, were used in the linkage. Information derived from the HTSF and used during linkage is removed from the final analytical file.

Table 1
Canadian Community Health Survey respondents eligible for record linkage to Canadian Mortality Database, by survey year, 2000/2001 to 2011

Canadian Community Health Survey year (cycle)		Agreed to share and link	
	number	number	%
Total	**720,426**	**614,774**	**85.3**
2000/2001 (1.1)	131,535	117,837	89.6
2003/2004 (2.1)	135,573	112,851	83.2
2005/2006 (3.1)	132,947	113,880	85.7
2007	65,946	57,083	86.6
2008	66,013	55,592	84.2
2009	61,679	52,475	85.1
2010	63,191	52,198	82.6
2011	63,542	52,858	83.2

Source: 2000/2001 to 2011 Canadian Community Health Survey.

Linkage

Linkage involved three steps: 1) data preparation; 2) record linkage; and 3) quality assessment. The linkage was approved by Statistics Canada's Executive Management Board.[20] Use of the linked data is governed by the Directive on Record Linkage.[21]

Data preparation

The following variables, which were chosen based on commonality among datasets, data quality and discriminatory power, were used for the linkage: given name, last name, date of birth, sex and postal code. Invalid responses for any of the variables were set to missing.

The rate of missing among the linkage variables was typically less than 3%. The exception was last names captured in the CCHS, where an average of 10% of respondents had missing information across survey cycles. The percentage of missing last names ranged from 4.1% in 2000/2001 to 17.1% in 2011. With linkage to the HTSF, the percentage of CCHS respondents with no last name was reduced to 2%. To reduce missed links due to misspelling, all names (given and last) were converted to their phonetic forms using the New York State Individual Intelligence System.

The discriminatory power of the linkage variables was assessed with Shannon entropy scores.[22] Higher levels of entropy reflect higher discriminatory power. Variables with larger numbers of distinct values (for example, last name) have higher discriminatory power and entropy levels, compared with variables with fewer distinct values (for example, sex).[23] Entropy levels among the linkage variables ranged from 0.43 for given names to 0.85 for postal codes (data not shown in tables). This information was employed in the record linkage strategy to assign initial linkage weights.

A single linkage cohort was created by concatenating the linkage variables of all CCHS cycles. This approach was more efficient and reduced the probability of false links, compared with linking each cycle separately. Additional details about the record linkage are available in an internal report.[24]

Record linkage

Record linkage of the CCHS and CMDB was accomplished in two steps. First, eligible CCHS respondents were linked to the HTSF to attach alternative postal codes and names. This information was used in the second step, where the CCHS was linked to the CMDB. All eligible CCHS respondents were linked to the CMDB, regardless of whether the record was linked to the HTSF.

The linkages were conducted in G-Link using probabilistic linkage methodology based on the Fellegi-Sunter theory of record linkage.[25] G-Link is a SAS-based generalized record linkage software developed at Statistics Canada to facilitate large-scale probabilistic record linkage.[26]

The CCHS was linked to the HTSF using the following variables: *given and last names, date of birth, postal code* and *sex*. The CCHS was linked to the CMDB using the same variables, which were available in both datasets. For records that linked to the HSTF, the following variables were also used: *alternative postal codes, alternative names (for example, maiden name, father's name), death date* and *interview date*. The availability of alternative postal codes in the HTSF has been found to improve linkage rates, particularly over time.[27]

Table 2
Number and percentage of Canadian Community Health Survey (CCHS) respondents who linked to Historical Tax Summary File (HTSF) and to Canadian Mortality Database (CMDB), by survey year, 2000/2001 to 2011

Canadian Community Health Survey year	Linked to HTSF				Linked to CMDB				Linked to CMDB (CCHS respondents who linked to HTSF)			
	Unweighted		Weighted		Unweighted		Weighted		Unweighted		Weighted	
	number	%	number	%	number	%	number	%	number	%	number	%
Total	**517,725**	**84.2**	**23,015,391**	**82.8**	**32,889**	**5.3**	.	.	**31,274**	**6.0**	.	.
2000/2001 (1.1)	101,495	86.1	22,063,102	85.5	11,768	10.0	1,979,060	7.7	11,187	11.0	1,869,609	8.5
2003/2004 (2.1)	92,881	82.3	21,302,890	80.2	8,672	7.7	1,399,526	5.3	8,222	8.9	1,327,901	6.2
2005/2006 (3.1)	98,590	86.6	23,190,360	85.5	6,685	5.9	1,092,021	4.0	6,345	6.4	1,033,059	4.5
2007	49,096	86.0	23,725,215	85.2	2,302	4.0	763,537	2.7	2,211	4.5	723,362	3.0
2008	47,046	84.6	23,696,053	84.0	1,689	3.0	552,636	2.0	1,620	3.4	534,562	2.3
2009	43,723	83.3	23,483,307	82.2	992	1.9	316,691	1.1	951	2.2	302,384	1.3
2010	42,962	82.3	23,534,665	81.5	622	1.2	220,092	0.8	586	1.4	199,831	0.8
2011	41,932	79.3	23,127,535	79.2	159	0.3	63,076	0.2	152	0.4	57,034	0.2

. not available for any reference period
Sources: 2000/2001 to 2011 Canadian Community Health Survey; 1996 to 2011 Historical Tax Summary File; 2000 to 2011 Canadian Mortality Database.

Linkage rates were calculated for both the CCHS-to-HTSF linkage and the CCHS-to-CMDB linkage. Rates represent the number of linked CCHS respondents divided by the number of CCHS respondents eligible for linkage (unweighted and weighted).

Quality assessment

Internal and external validation was conducted to evaluate the accuracy of the linkage process and the fitness of the data for use in analysis.

Table 3
Number and percentage of Canadian Community Health Survey (CCHS) respondents aged 12 or older who linked to Canadian Mortality Database (CMDB), by province/territory, sex and age group, 2000/2001 to 2011

Province/Territory, sex and age group	CCHS respondents who linked to CMDB			
	Unweighted		Weighted	
	number	%	number	%
Total	**32,889**	**5.3**	..	..
Province/Territory				
Newfoundland	950	5.0	15,374	3.4
Prince Edward Island	746	6.6	4,279	3.6
Nova Scotia	1,622	6.6	30,644	3.8
New Brunswick	1,372	5.6	21,683	3.4
Quebec	6,068	4.9	198,841	3.0
Ontario	10,382	5.3	291,177	2.7
Manitoba	2,228	6.2	32,065	3.4
Saskatchewan	2,413	6.6	28,913	3.6
Alberta	2,638	4.5	65,655	2.3
British Columbia	4,144	5.7	108,248	2.9
Yukon	122	3.0	532	1.9
Northwest Territories	118	2.4	587	1.7
Nunavut	86	2.5	327	2.0
Sex				
Male	16,439	5.8	429,275	3.1
Female	16,450	5.0	369,054	2.6
Age group				
12 to 24	310	0.3	9,138	0.2
25 to 34	356	0.4	10,804	0.2
35 to 44	932	1.0	29,068	0.6
45 to 54	2,242	2.4	73,536	1.5
55 to 64	4,482	4.8	122,125	3.2
65 to 74	8,233	11.4	191,224	8.0
75 or older	16,334	27.0	362,432	20.5

.. not available for a specific reference period
Sources: 2000/2001 to 2011 Canadian Community Health Survey; 2000 to 2011 Canadian Mortality Database.

Internal validation

False positive and false negative rates were calculated. Manual review (blind) was conducted on a random sample of 1,000 pairs uniformly selected across eight strata of linkage weights representing the full range of linkage weights above the threshold (that is, links), but only a limited range of weights just below the threshold (non-links). This approach is recommended given the large number of non-links generated in the initial creation of pairs, which, if fully represented, would result in an artificially low false negative rate.[28]

Three reviewers independently conducted a manual review of the same information used in the linkage (names, date of birth, postal code and sex). Additional information available for the manual review included date of death from the HTSF for CCHS respondents who linked to the tax data and had died. Each pair was then assigned a link or non-link status.

Results of the manual review were compared with G-Link results to calculate stratum-specific estimates of false positive and false negative links.[26,29] Global false positive and false negative rates were estimated using a weighting function based on the distribution of all possible pairs. Although a uniform sample allocation was used to select pairs from each weight stratum, the sampling fractions differed because some strata contained more pairs than did others. About 75% of pairs belonged to the first stratum, where linkage weights were

lowest; fewer than 1% of pairs belonged to the last stratum (stratum 8), where linkage weights were highest.

External validation

Linkage rates to the CMDB were calculated by basic respondent characteristics (for instance, age and sex). It was anticipated that patterns of linkage rates would reflect the differential risk of death among groups at higher risk, such as males and older adults.[30,31]

The linked data were also externally validated by comparing the distribution of deaths identified in the linked cohort with the distribution of deaths derived from the CMDB alone for people aged 12 or older. Distributions by geography, age, sex and major cause of death were examined. Age-standardized mortality rates (ASMRs) were calculated by calendar year for each CCHS cycle and compared with official mortality statistics.[32]

Cox proportion hazards survival models (age-adjusted) were used to assess the impact of several behavioural risk factors on mortality. The analysis pertained to respondents aged 20 or older from the 2003 and 2005 CCHS linked data, with up to eight years of mortality follow-up. Based on the external validation, the first year of follow-up was excluded from the analysis.

Table 4
Unweighted and weighted percentage distribution of deaths among those aged 12 or older identified in Canadian Community Health Survey (CCHS) linked cohort, compared with Canadian Mortality Database (CMDB), by province/territory, sex, age group and cause of death, 2000/2001 to 2011

	CCHS respondents who linked to CMDB		
	Unweighted	Weighted	CMDB
	%		
Province/Territory			
Newfoundland	2.9	1.9	1.9
Prince Edward Island	2.3	0.5	0.5
Nova Scotia	4.9	3.8	3.6
New Brunswick	4.2	2.7	2.7
Quebec	18.4	24.9	24.3
Ontario	31.6	36.5	37.0
Manitoba	6.8	4.0	4.3
Saskatchewan	7.3	3.6	3.9
Alberta	8.0	8.2	8.4
British Columbia	12.6	13.6	13.2
Yukon	0.4	0.1	0.1
Northwest Territories	0.4	0.1	0.1
Nunavut	0.3	0.0	0.0
Sex			
Male	50.0	53.8	50.5
Female	50.0	46.2	49.5
Age group			
12 to 24	0.9	1.1	1.2
25 to 34	1.1	1.4	1.2
35 to 44	2.8	3.6	2.6
45 to 54	6.8	9.2	6.0
55 to 64	13.6	15.3	10.6
65 to 74	25.0	24.0	17.7
75 or older	49.0	45.4	60.7
Cause of death			
Diseases of blood and blood-forming organs and certain disorders involving immune system	0.4	0.5	0.4
Cancers	33.9	36.1	30.2
Diseases of circulatory system	31.0	28.9	31.2
Diseases of digestive system	4.0	3.8	3.9
Endocrine, nutritional and metabolic diseases	4.1	3.9	4.2
External causes of morbidity and mortality	5.5	5.7	6.4
Diseases of genitourinary system	2.3	2.4	2.2
Certain infectious and parasitic diseases	1.7	2.0	1.8
Congenital malformations, deformations and chromosomal abnormalities	0.1	0.2	0.2
Mental and behavioural disorders	3.0	3.0	4.0
Diseases of musculoskeleteral system and connective tissue	0.7	0.7	0.7
Diseases of nervous system	3.4	3.7	4.8
Diseases of respiratory system	8.6	8.3	8.6
Diseases of skin and subcutaneous tissue	0.2	0.1	0.2
Symptoms, signs and abnormal clinical and laboratory findings and chromosomal abnormalities	1.0	0.9	1.1
Other causes	0.0	0.0	0.0

Sources: 2000/2001 Canadian Community Health Survey; Canadian Mortality Database.

Covariates

Results of the linkage are presented by province and territory, sex and age at the time of the CCHS interview. Results of the survival analysis are reported for five risk factors as defined in the CCHS: smoking status, body mass index (BMI), alcohol consumption, physical activity, and fruit and vegetable consumption. Self-reported BMI values were adjusted with the recommended correction factor.[33] Analyses were conducted using the special sampling weights to adjust for the exclusion of CCHS respondents who did not agree to share and link their data.

Respecting respondent privacy

Statistics Canada ensures respondent privacy during the linkage process and subsequent use of the linked files. Only employees directly involved in the process had access to the unique identifying information required for linkage (such as names); these individuals did not access health-related information. Once the data linkage was completed, an analytical file was created from which identifying information was removed. This de-identified file could then be accessed by analysts for validation and analysis.

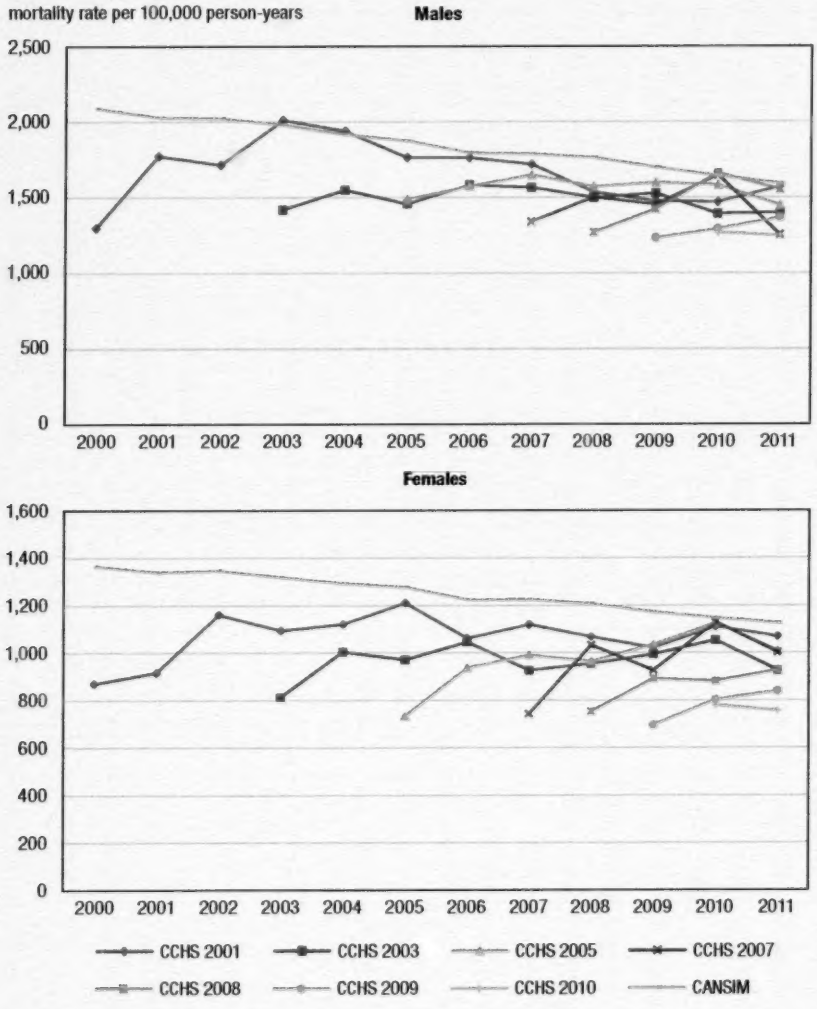
Results

Linkage

A large majority (84.2%) of eligible CCHS respondents linked to the HTSF, with rates ranging across survey years from 79.3% to 86.6% (Table 2). Linkage rates were low (78%) for British Columbia and Nunavut compared with other regions (data not shown in tables). As well, linkage rates were low among respondents younger than 25 (74.3%) or older than 74 (79.4%), and high among 25- to 34-year-olds (89%) (data not shown in tables). These patterns by region and age persisted over the survey years (data not shown).

Overall, 5.3% of eligible CCHS respondents linked to the CMDB (Table 2). The linkage rate to the CMDB was slightly higher among CCHS respondents who linked to the HTSF (6.0%). As expected, given the shorter follow-up period, linkage rates were lower for more recent survey years. Respondents in the territories were less likely to link to the CMDB (2.4% to 3.0%) than were those in other regions.

Figure 1
Age-standardized mortality rates per 100,000 person-years, by sex and survey year, Canadian Community Health Survey (CCHS)-Canadian Mortality Database linked data and CANSIM, 2000 to 2011



Source: 2000/2001 to 2011 Canadian Community Health Survey; CANSIM table 102-0504.

Quality assessment

The false positive and false negative rates were 0.04% and 2.43%, respectively (data not shown in tables). As expected, linkage rates (unweighted) to the CMDB were higher among males (5.8%) than females (5.0%) and rose with age, with the highest rates among those aged 75 or older (27.0%) (Table 3). This pattern of linkage rates was similar across survey years (data not shown in tables).

Geographic distributions of deaths in the linked CCHS-CMDB data were similar to those derived from the CMDB alone. The majority of deaths occurred in the largest provinces (Ontario, Quebec and British Columbia) (Table 4). While overall patterns persisted, the weighted distribution of deaths across provinces and territories resembled results from the CMDB more closely than did the unweighted distributions. Differences were particularly evident for the Atlantic Provinces.

The distribution of deaths by age group revealed the expected increase at older ages (Table 4). However, as a percentage of all deaths, 49% were among people aged 75 or older in the linked data, compared with 61% in the CMDB data. The difference largely reflects deaths among the institutionalized population in the CMDB, which are not represented in the linked CCHS-CMDB data. When only deaths captured in the CMDB among people aged 12 through 74 were considered, the distributions were similar (data not shown). The distribution of deaths by cause derived from the linked data was similar to the distribution in the CMDB (Table 4).

Annual age-standardized mortality rates (ASMRs) derived from the linked CCHS-CMDB data were generally lower than those derived from official mortality rates (Figure 1). ASMRs were lowest in the first year of follow-up for all CCHS years, indicating a potential "healthy

respondent bias," which supports the decision to exclude the first year of follow-up from the survival models.

Survival analysis

The survival analysis generally reveals higher hazard ratios (HRs) for mortality among groups at greater risk (Table 5). Heavy smokers (males = 2.36; females = 2.91) and light smokers (males = 1.92; females = 1.81) had elevated HRs compared with non-smokers. The mortality risk associated with BMI was U-shaped, with elevated HRs for individuals in the underweight class (males = 1.77; females = 1.50) or in obese class II or above (males = 1.51; females = 1.20). HRs for former drinkers (males = 1.65; females = 1.56) were high, compared with moderate drinkers. Reporting no physical activity or less than 30 minutes a day was associated with greater mortality risks, compared with reporting 60 or more minutes a day. Finally, eating fewer than five servings of fruit and vegetables per day was associated with elevated mortality risks, compared with consuming more than five servings; findings for no servings were not significant, likely because of small sample sizes.

Discussion

Overall, 5.3% of eligible Canadian Community Health Survey respondents were linked to a death record in the Canadian Mortality Database; false positive and false negative rates were 0.04% and 2.43%, respectively. Use of the Historical Tax Summary File yielded slightly higher linkage rates to the CMDB. External validation revealed patterns in mortality rates in the linked CCHS-CMDB data that were comparable to nationally reported estimates. These patterns attest to the quality of the linked data and the suitability of these data for mortality research at the population level.

Table 5
Age-adjusted hazard ratios for mortality associated with selected health behaviours, by sex, respondents aged 20 or older to 2003 and 2005 Canadian Community Health Surveys linked to Canadian Mortality Database

	Men			Women		
		95% confidence interval			95% confidence interval	
Health behaviour	Hazard ratio	from	to	Hazard ratio	from	to
Smoking						
Non-smoker[§]	1.00	...	...	1.00	...	...
Light smoker	1.92*	1.51	2.33	1.81*	1.52	2.11
Heavy smoker	2.36*	1.84	2.89	2.91*	1.92	3.91
Former smoker	1.23*	1.05	1.40	1.31*	1.16	1.46
Body Mass Index (BMI), with correction						
Underweight (less than 18.5)	1.77*	1.06	2.47	1.50*	1.16	1.85
Normal weight (18.5 to less than 25.0)[§]	1.00	...	...	1.00	...	...
Overweight (25.0 to less than 30.0)	0.87*	0.79	0.95	0.86*	0.78	0.94
Obese I (30.0 to less than 35.0)	0.96	0.85	1.07	0.91	0.80	1.02
Obese II (35.0 or more)	1.51*	1.25	1.76	1.20*	1.00	1.40
Alcohol consumption						
Light or non-drinker	1.20*	1.09	1.31	1.15*	1.01	1.29
Moderate drinker[§]	1.00	...	...	1.00	...	...
Heavy drinker	1.35*	1.06	1.63	2.29	0.41	4.16
Former drinker	1.65*	1.46	1.83	1.56*	1.36	1.75
Minutes of physical activity per day						
None	1.89*	1.61	2.16	2.04*	1.65	2.43
Less than 30	1.23*	1.08	1.38	1.22*	1.03	1.41
30 to 60	0.99	0.86	1.12	1.10	0.92	1.28
More than 60[§]	1.00	...	...	1.00	...	...
Fruit and vegetable servings per day						
None	1.93	0.49	3.36	1.71	0.57	2.85
Less than 2	1.52*	1.30	1.74	1.82*	1.51	2.13
2 to 5	1.18*	1.03	1.32	1.18*	1.05	1.32
More than 5[§]	1.00	...	...	1.00	...	...

... not applicable
* siginficantly different from reference category ($p < 0.05$)
[§] reference category
Sources: 2003 and 2005 Canadian Community Health Survey; 2004 to 2011 Canadian Mortality Database.

What is already known on this subject?

- Vital statistics registries and national health surveys are complementary sources of data.
- When administrative and health survey data are combined through record linkage, relationships between social determinants and health outcomes such as death can be analyzed in more depth.
- In Canada, mortality data have been linked to census results to examine differences in the risk of death across socioeconomic groups and to calculate mortality rates for subpopulations such as immigrants and Indigenous people.
- An Ontario study used linked health survey and mortality data to explore associations with smoking, diet, physical activity and stress.

What does this study add?

- Data from the 2000/2001 to 2011 Canadian Community Health Survey (CCHS) have been linked to the Canadian Mortality Database (CMDB).
- Overall, 5.3% of CCHS respondents were linked to a death record in the CMDB.
- The similarity of patterns in mortality rates based on the linked data and those in nationally reported estimates indicates that the linked data are suitable for studying mortality at the population level in Canada.

Age-standardized mortality rates derived from the linked data were consistently lower in the first and second years of follow-up, compared with national official mortality estimates. This was anticipated, given that the official estimates are based on the entire population, which includes institutionalized individuals, whereas rates estimated from the linked CCHS-CMDB data represent only the household population.

Low ASMRs in the first two years of follow-up may also reflect a "healthy respondent" bias. CCHS respondents may have a more favourable morbidity and mortality profile than non-respondents. People who are ill or near death may be less likely to respond to a survey. Evidence from a Scottish health survey linked to mortality data revealed higher mortality rates among non-respondents (55 or older) than respondents.[34] Similarly, in Finland, linked survey and mortality data showed higher mortality rates among non-respondents than respondents to the annual health survey.[35]

Differences in mortality rates could be influenced by the demographic and socioeconomic status of survey respondents versus non-respondents.[35,36] In addition, non-respondents may be at higher risk for alcohol-, drug- and smoking-related mortality than are respondents.[37] This suggests that the impact of survey non-respondents may be to underestimate overall mortality and the influence of risk behaviours.

Consistent with international findings,[18,38-41] the survival analysis based on the linked CCHS-CMDB data revealed elevated hazard ratios for mortality among smokers, people with very low or very high BMI, those who reported fewer than five servings of fruit and vegetables per day, and those who reported less than 30 minutes of physical activity per day.

For more than half a century, large epidemiological studies have documented risks of premature mortality related to smoking.[42] A U-shaped curve for BMI and mortality has been shown in both Canadian and American research.[43] The pattern for alcohol consumption and mortality aligns with an international meta-analysis,[44] and the elevated risk among former drinkers compared with moderate drinkers supports the "sick quitter" effect (individuals stop drinking because of disease onset).[45]

Recent research[46,47] has indicated that, paradoxically, some traditional risk factors might actually reverse at older ages, hypotheses that future analyses might test with linked CCHS-CMDB data. Other social determinants of health, such as social interactions and support that may contribute to mortality risk, particularly among the elderly,[48-51] could be investigated using the linked data.

Limitations

A degree of caution is warranted in using the linked CCHS-CMDB data. The low mortality rates in the first year of follow-up suggest a "healthy respondent" bias. Consequently, it is recommended that at least the first year of follow-up be excluded from mortality estimates. As well, although the CCHS is cross-sectional, individuals may be represented more than once within and across cycles. These respondents were not removed, but rather, identified and flagged in order to link them to the same death record. However, this affects only analyses requiring more than one survey year.

Conclusions

The CCHS can be linked to the CMDB to examine associations between health risk factors and mortality. The linkage constitutes a resource for research on mortality outcomes. These data can be used to refine estimates of associations between risk behaviours and mortality in predictive models of all-cause and cause-specific mortality. ■

Acknowledgments

The authors acknowledge funding from Health Canada, Institute for Clinical and Evaluative Sciences/Ottawa Hospital Research Institute and McGill University to support the linkage project.

References

1. Holman CD, Bass AJ, Rosman DL, et al. A decade of data linkage in Western Australia: Strategic design, applications and benefits of the WA data linkage system. *Australian Health Review* 2008; 32(4): 766-77.
2. Winkler WE. Record linkage. *Sample Surveys: Design, Methods and Applications* 2009; 29A: 351-80.
3. Rotermann M, Sanmartin G, Trudeau R, et al. Two approaches to linking census and hospital data. *Health Reports* 2014; 25(10): 3-14.
4. Wilkins K, Shields M, Rotermann M. Smokers' use of acute-care hospitals—A prospective study. *Health Reports* 2009; 20(4): 75-84.
5. Fransoo R, Martens P, Prior H, et al. *Adult Obesity in Manitoba: Prevalence, Associations, and Outcomes*. Winnipeg, Manitoba: Manitoba Centre for Health Policy, 2011.
6. Davidsen M, Kjøller M, Helweg-Larsen K. The Danish National Cohort Study (DANCOS). *Scandinavian Journal of Public Health* 2011; 39(7 Suppl): 131-5.
7. Teckle P, Hannaford P, Sutton M. Is the health of people living in rural areas different from those in cities? Evidence from routine data linked with the Scottish Health Survey. *BMC Health Services Research* 2012; 12: 43.
8. Hamer M, Batty GD, Stamatakis E, Kivimaki M. Comparison of risk factors for fatal stroke and ischemic heart disease: a prospective follow up of the health survey for England. *Atherosclerosis* 2011; 219(2): 807-10.
9. Wilkins R, Tjepkema M, Mustard C, et al. The Canadian census mortality follow-up study, 1991 through 2001. *Health Reports* 2008; 19(3): 25-43.
10. Hwang SW, Wilkins R, Tjepkema M, et al. Mortality among residents of shelters, rooming houses, and hotels in Canada: 11-year follow-up study. *British Medical Journal* 2009; 339: b4036. doi:101136/bmj.b4036.
11. Tjepkema M, Wilkins R. Remaining life expectancy at age 25 and probability of survival to age 75, by socio-economic status and Aboriginal ancestry. *Health Reports* 2011; 22(4): 31-6.
12. Ng E. The healthy immigrant effect and mortality rates. *Health Reports* 2011; 22(4): 24-30.
13. Tjepkema M, Wilkins R, Senécal S, et al. Potential years of life lost at ages 25 to 74 among Métis and non-Status Indians, 1991 to 2001. *Health Reports* 2011; 22(1): 37-46.
14. Dankovchik J, Hoopes MJ, Warren-Mears, Knaster E. Disparities in life expectancy of Pacific Northwest American Indians and Alaska natives: analysis of linkage-corrected life tables. *Public Health Reports* 2015; 130(1): 71-80.
15. Kariminia A, Butler TG, Corben SP, et al. Extreme cause-specific mortality in a cohort of adult prisoners–1988 to 2002: a data-linkage study. *International Journal of Epidemiology* 2007; 36: 310-6.
16. Kinner SA, Degenhardt L, Coffey C, et al. Substance use and risk of death in young offenders: A prospective data linkage study. *Drug and Alcohol Review* 2015; 34: 46-50.
17. Nambiar D, Weir A, Aspinall EJ, et al. Mortality and cause of death in a cohort of people who had ever injected drugs in Glasgow: 1982-2012. *Drug and Alcohol Dependence* 2015; 1470: 215-21.
18. Manuel DG, Perez R, Bennett C, et al. *Seven More Years: The Impact of Smoking, Alcohol, Diet, Physical Activity, and Stress on Health and Life Expectancy in Ontario*. An ICES/PHO Report. Toronto: Clinical Evaluative Sciences and Public Health Ontario, 2012.
19. Statistics Canada. *Canadian Community Health Survey*. Available at: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3226&lang=en&db=imdb&adm=8&dis=2#b4
20. Statistics Canada. *Approved Record Linkages*. Available at: http://www.statcan.gc.ca/record-enregistrement/summ-somm-eng.htm
21. Statistics Canada. *Directive on Record Linkage*. Available at: http://www.statcan.gc.ca/record-enregistrement/policy4-1-politique4-1-eng.htm
22. Shannon CE. A mathematical theory of communication. *The Bell System Technical Journal* 1948; 27: 379-423, 623-56.
23. Newcombe HB. Record linking: The design of efficient systems for linking records into individual and family histories. *American Journal of Human Genetics* 1967; 19: 335-59.
24. Statistics Canada. *Summary Report of the Canadian Community Health Survey (CCHS) – Canadian Mortality Database (CMDB) Record Linkage Study* (internal document). Ottawa: Statistics Canada, Health Statistics Division, 2015.
25. Fellegi IP, Sunter AB. A theory for record linkage. *Journal of the American Statistical Association* 1969; 64(328): 1183-1210.
26. Statistics Canada. *G-Link Version 3.0 User Guide* (internal document).
27. Rotermann M, Sanmartin C, Trudeau R, St-Jean H. Linking 2006 Census and hospital data in Canada. *Health Reports* 2015; 26(10): 10-20.
28. Agency for Healthcare Research and Quality. *Linking Data for Health Services Research: A Framework and Instructional Guide*. AHRQ Publication No. 14-EHC033_EF. Rockville, Maryland: Department of Health and Human Services, 2014.
29. Dohoo I, Martin W, Stryhn H. *Methods in Epidemiologic Research*. Charlottetown, Prince Edward Island: VER Inc., 2012.
30. Statistics Canada. *Trends in Mortality Rates 2000-2011*. Available at: http://www.statcan.gc.ca/pub/82-625-x/2014001/article/11897-eng.htm
31. Statistics Canada. Deaths, by age and sex, Canada. CANSIM Table 102-0503. Available at: http://www5.statcan.gc.ca/cansim/a26
32. Statistics Canada. Deaths and mortality rates, by age group and sex, Canada, provinces and territories. CANSIM Table 102-0504. Available at: http://www5.statcan.gc.ca/cansim/pick-choisir?lang=eng&p2=33&id=1020504
33. Shields M, Connor Gorber S, Janssen I, Tremblay MS. Bias in self-reported estimates of obesity in Canadian health surveys: An update on correction equations for adults. *Health Reports* 2011; 22 (3): 35-45.
34. Cohen G, Dugg JC. Are nonrespondents to health surveys less healthy than respondents? *Journal of Official Statistics* 2002; 18(1): 13-23.
35. Tolonen H, Laatikainen T, Helakorpi S, et al. Marital status, educational level and household income explain part of the excess mortality of survey non-respondents. *European Journal of Epidemiology* 2010; 25(2): 69-76. doi: 10.1007/s10654-009-9389-9. Epub 2009 Sep 25.
36. Laaksonen M, Talala K, Martelin T, et al. Health behaviours as explanations for educational level differences in cardiovascular and all-cause mortality: a follow-up of 60,000 men and women over 23 years. *European Journal of Public Health* 2007; 18(1): 38-43.
37. Christensen AI, Ekholm O, Gray L, et al. What is wrong with non-respondents? Alcohol, drug- and smoking related mortality and morbidity in a 12-year follow up study of respondents and non-respondents in the Danish Health and Morbidity Survey. *Addiction* 2015; 110(9): 1505-12.

38. Ford ES, Zhao G, Tsai J, Li C. Low-risk lifestyle behaviors and all-cause mortality: Findings from the National Health and Nutrition Examination Survey III Mortality Study. *American Journal of Public Health* 2011; 101(10): 1922.

39. Ford ES, Bergmann MM, Boeing H, et al. Healthy lifestyle behaviors and all-cause mortality among adults in the United States. *Preventive Medicine* 2012; 55(1): 23-7.

40. Behrens G, Fischer B, Kohler S, et al. Healthy lifestyle behaviors and decreased risk of mortality in a large prospective study of U.S. women and men. *European Journal of Epidemiology* 2013; 28(5): 361-72.

41. Oyebode O, Gordon-Dseagu V, Walker A, Mindell JS. Fruit and vegetable consumption and all-cause, cancer and CVD mortality: Analysis of Health Survey for England data. *Journal of Epidemiology and Community Health* 2014; 68(9): 856-62.

42. Thun MJ. Public Health Classics: When truth is unwelcome: The first reports on smoking and lung cancer. *Bulletin of the World Health Organization* 2005; 83(2): 144-5.

43. Flegal KM, Kit BK, Orpana H, Graubard BI. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. *Journal of the American Medical Association* 2013; 309(1): 71-82.

44. Ronksley PE, Brien SE, Turner BJ, et al. Association of alcohol consumption with selected cardiovascular disease outcomes: A systematic review and meta-analysis. *British Medical Journal* 2011; 342 (7795)

45. Shaper AG. Alcohol consumption decreases with the development of disease. *Addiction* 2011; 106(5): 1023-5.

46. Peter RS, Mayer B, Concin H. The effect of age on the shape of the BMI-mortality relation and BMI associated with minimum all-cause mortality in a large Austrian cohort. *International Journal of Obesity* 2015; 39(3): 530-4.

47. Ahmadi S-F, Streja E, Zahmatkesh G. Reverse epidemiology of traditional cardiovascular risk factors in the geriatric population. *Journal of the American Medical Directors Association* 2015; 16(11): 933-9.

48. Centers for Disease Control. *Heat Waves 2009.* [cited August 1 2014]. Available at: http://www.cdc.gov/climateandhealth/effects/heat.htm. Accessed August 1, 2014.

49. Kleninberg E. Dying alone: The social production of urban isolation. *Ethnography* 2001; 2(4): 501-31.

50. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. *PLoS Medicine* 2010; 7(7): 1000316.

51. Stringhini S, Berkman L, Dugravot A, et al. Socioeconomic status, structural and functional measures of social support, and mortality: The British Whitehall II Cohort Study, 1985-2009. *American Journal of Epidemiology* 2012; 17 (12): 1275-83.